Det egna projektet steg 1
Uppgift:
Förutspå utfall i Premier League-matcher

Målet är att träna en modell att förutse utfall i Premier League-matcher baserat på matchdata flera år bakåt i tiden.

Datat ska samlas in med ett scraping-script från en site som erbjuder historisk data gratis, alternativt från ett befintligt dataset på kaggle.com.
Datasetet ser ut enligt följande:

```
,date,time,comp,round,day,venue,result,gf,ga,opponent,xg,xga,poss,attendance,captai
n,formation,referee,match report,notes,sh,sot,dist,fk,pk,pkatt,season,team
0,2023-08-13,16:30,Premier League,Matchweek
1,Sun,Away,D,1.0,1.0,Chelsea,1.3,1.4,35.0,40096.0,Virgil van Dijk,4-3-3,Anthony
Taylor,Match Report,,13.0,1.0,17.8,0.0,0,0,2024,Liverpool
1,2023-08-19,15:00,Premier League,Matchweek
2,Sat,Home,W,3.0,1.0,Bournemouth,3.0,1.3,64.0,53145.0,Virgil van Dijk,4-3-3,Thomas
Bramall,Match Report,,25.0,9.0,16.8,1.0,0,1,2024,Liverpool
```

Datat kommer vara komplett och innehålla några säsongers matchsdata, och kommer med flera olika datatyper, där de som ska användas kommer behöva göras om till integers eller datetime.

Eftersom jag ska försöka förutse ett begränsat antal utfall är det ett klassificeringsproblem det handlar om.

Tanken är att testa modellen med RandomForestClassifier och utvärdera den med ett par metriker i sklearn för att se vilken som ger bäst resultat.